Finding Nemo

Harsh Mehta
Rohitashwa Chakraborty
Karthick Ramasubramanian

Executive Summary

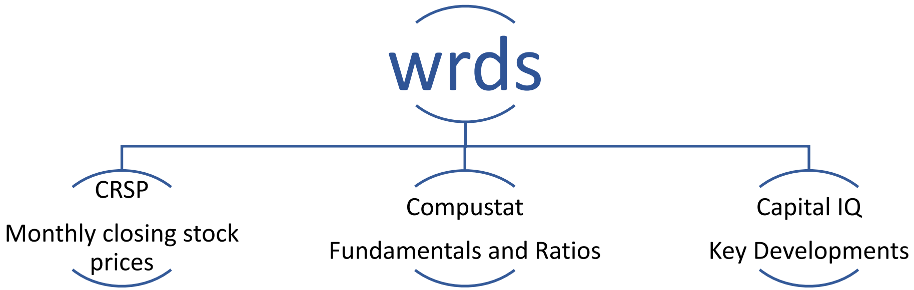
Using industry fundamental data, text data from news articles, our team has attempted to predict the direction of change of Net Income using Machine Learning.

We would use the results to build quarterly balanced portfolios and test it against Market and Fama French 3 factor Portfolios.

Objective

- Forecast the direction change in Net Income
- Without Normalizing using Shares Outstanding or Revenue we believe the data would be less granular and hence more accurate thereby granting our model more predictive power
- Our final objective is to build portfolios with statistically significant Alphas

Data Sources



- We are using data starting from 2001 to 2020
- The fundamental company data has been obtained from Compustat.
- The News data obtained from Capital IQ (Compustat) includes:
    - Category of the Article - Sales, Acquisitions, Stock Splits etc..
    - The header
    - A brief description of the news article
- Stock prices were pulled from CRSP
- Market Data and the Fama French 3 factor model data has been obtained from Kenneth R French Data Library

Approach

Data Preprocessing

- Exploratory Data Analysis
- Dropping rows and columns with missing values
- Preparing target variable

Feature Engineering

- Adding % change columns
- Cleaning the unstructured data in the key events file.
- Mapping event headlines and implementing sentiment analysis on the cleaned vectorized data from key events.

Model Building

- Univariate logistic regressions to find significant variables
- Multivariate logistic regression on subset of variables
- Train on previous 4 quarters and predict change in direction in NI for next quarter

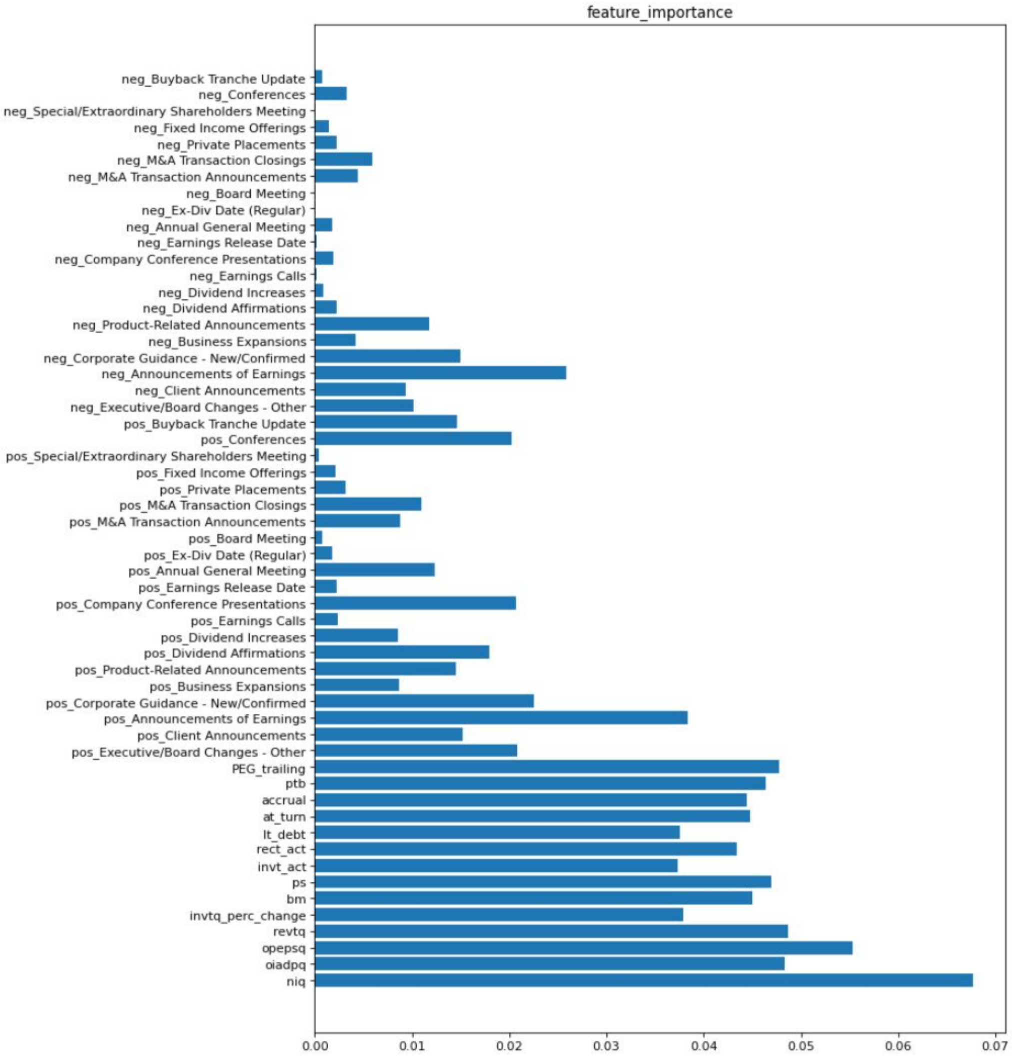
Portfolio Construction

- Equal Weighted portfolios
- Different probability cutoffs considered
- Rebalanced quarterly
- Regressed returns against fama-French 3 factor model

Variables Used

```
Index(['cusip', 'dir', 'public_date', 'niq', 'oiadpq', 'opepsq', 'revtq',
       'invtq_perc_change', 'bm', 'ps', 'invt_act', 'rect_act', 'lt_debt',
       'at_turn', 'accrual', 'ptb', 'PEG_trailing',
       'pos_Executive/Board Changes - Other', 'pos_Client Announcements',
       'pos_Announcements of Earnings',
       'pos_Corporate Guidance - New/Confirmed', 'pos_Business Expansions',
       'pos_Product-Related Announcements', 'pos_Dividend Affirmations',
       'pos_Dividend Increases', 'pos_Earnings Calls',
       'pos_Company Conference Presentations', 'pos_Earnings Release Date',
       'pos_Annual General Meeting', 'pos_Ex-Div Date (Regular)',
       'pos_Board Meeting', 'pos_M&A Transaction Announcements',
       'pos_M&A Transaction Closings', 'pos_Private Placements',
       'pos_Fixed Income Offerings',
       'pos_Special/Extraordinary Shareholders Meeting', 'pos_Conferences',
       'pos_Buyback Tranche Update', 'neg_Executive/Board Changes - Other',
       'neg_Client Announcements', 'neg_Announcements of Earnings',
       'neg_Corporate Guidance - New/Confirmed', 'neg_Business Expansions',
       'neg_Product-Related Announcements', 'neg_Dividend Affirmations',
       'neg_Dividend Increases', 'neg_Earnings Calls',
       'neg_Company Conference Presentations', 'neg_Earnings Release Date',
       'neg_Annual General Meeting', 'neg_Ex-Div Date (Regular)',
       'neg_Board Meeting', 'neg_M&A Transaction Announcements',
       'neg_M&A Transaction Closings', 'neg_Private Placements',
       'neg_Fixed Income Offerings',
       'neg_Special/Extraordinary Shareholders Meeting', 'neg_Conferences',
       'neg_Buyback Tranche Update'],
      dtype='object')
```

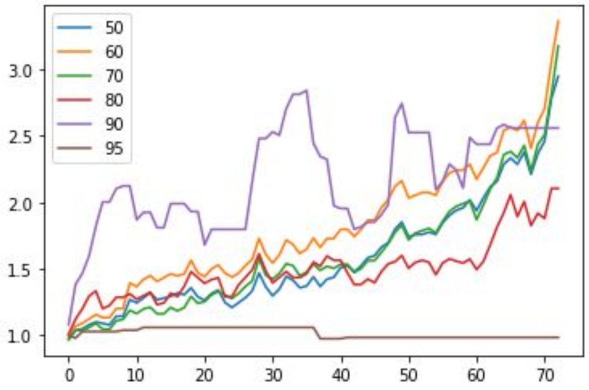
AUC = 0.67



feature_importance
neg_Buyback Tranche Update
neg_Conferences
neg_Special/Extraordinary Shareholders Meeting
neg_Fixed Income Offerings
neg_Private Placements
neg_M&A Transaction Closings
neg_M&A Transaction Announcements
neg_Board Meeting
neg_Ex-Div Date (Regular)
neg_Annual General Meeting
neg_Earnings Release Date
neg_Company Conference Presentations
neg_Earnings Calls
neg_Dividend Increases
neg_Dividend Affirmations
neg_Product-Related Announcements
neg_Business Expansions
neg_Corporate Guidance - New/Confirmed
neg_Announcements of Earnings
neg_Client Announcements
neg_Executive/Board Changes - Other
pos_Buyback Tranche Update
pos_Conferences
pos_Special/Extraordinary Shareholders Meeting
pos_Fixed Income Offerings
pos_Private Placements
pos_M&A Transaction Closings
pos_M&A Transaction Announcements
pos_Board Meeting
pos_Ex-Div Date (Regular)
pos_Annual General Meeting
pos_Earnings Release Date
pos_Company Conference Presentations
pos_Earnings Calls
pos_Dividend Increases
pos_Dividend Affirmations
pos_Product-Related Announcements
pos_Business Expansions
pos_Corporate Guidance - New/Confirmed
pos_Announcements of Earnings
pos_Client Announcements
pos_Executive/Board Changes - Other
PEG_trailing
ptb
accrual
at_turn
lt_debt
rect_act
invt_act
ps
bm
invtq_perc_change
revtq
opepsq
oiadpq
niq
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07

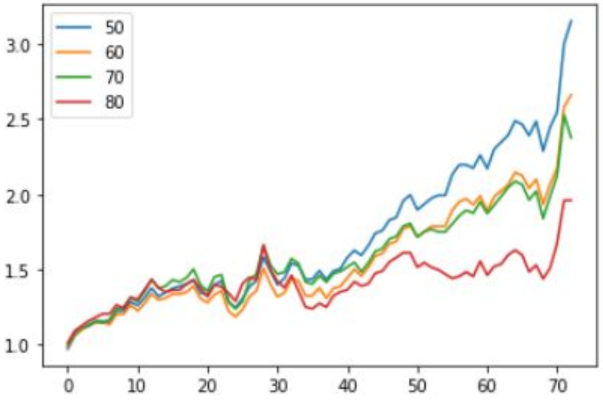
Only Structured Data

Long Only



	coef	std err	t	P>|t|	[0.025	0.975]
const	0.0173	0.007	2.432	0.018	0.003	0.032
Mkt-RF	-0.0009	0.002	-0.601	0.550	-0.004	0.002
SMB	0.0027	0.003	1.026	0.308	-0.003	0.008
HML	-0.0025	0.002	-1.249	0.216	-0.007	0.002
RF	-0.0090	0.043	-0.210	0.835	-0.095	0.077

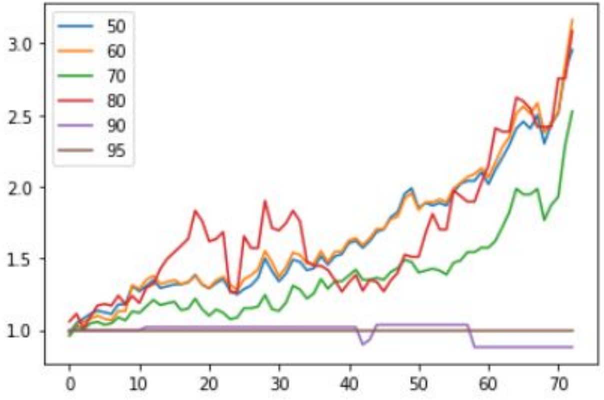
Long Short



	coef	std err	t	P>|t|	[0.025	0.975]
const	0.0156	0.007	2.123	0.037	0.001	0.030
Mkt-RF	0.0001	0.002	0.070	0.944	-0.003	0.003
SMB	0.0032	0.003	1.190	0.238	-0.002	0.009
HML	-0.0035	0.002	-1.655	0.103	-0.008	0.001
RF	-0.0315	0.044	-0.709	0.481	-0.120	0.057

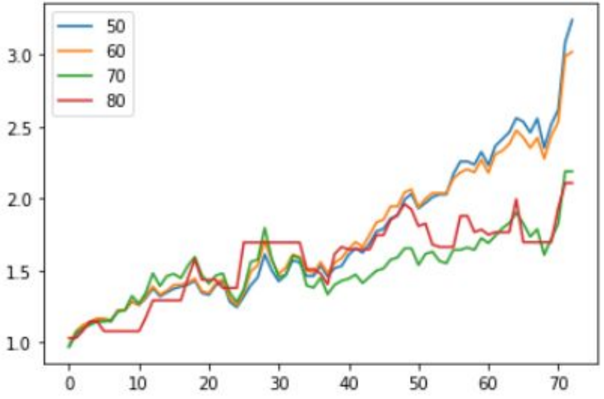
Structured and Unstructured Data

Long Only



	coef	std err	t	P>|t|	[0.025	0.975]
const	0.0175	0.007	2.406	0.019	0.003	0.032
Mkt-RF	0.0004	0.002	0.288	0.774	-0.003	0.003
SMB	0.0024	0.003	0.880	0.382	-0.003	0.008
HML	-0.0014	0.002	-0.695	0.489	-0.006	0.003
RF	-0.0224	0.044	-0.510	0.612	-0.110	0.065

Long Short



	coef	std err	t	P>|t|	[0.025	0.975]
const	0.0168	0.007	2.246	0.028	0.002	0.032
Mkt-RF	0.0003	0.002	0.180	0.858	-0.003	0.003
SMB	0.0032	0.003	1.149	0.255	-0.002	0.009
HML	-0.0031	0.002	-1.459	0.149	-0.007	0.001
RF	-0.0263	0.045	-0.581	0.563	-0.117	0.064

Next Steps

- Trying different transformations of the features(log) and adding lags
- Implementing NN and SVM and find a good balance between model complexity and explainability
- Analyzing richer unstructured data like Earnings Call Transcripts
- Fitting a ARIMA model to understand the predictive power of lagged effects

References

OU, Jane A, and Stephen H Penman. “Improving Earnings Predictions and Abnormal Returns with Machine Learning.” *https://www.sciencedirect.com/Science/Article/Abs/Pii/016541018990017 7*, Nov. 1989,

Hunt, Joshua. “Improving Earnings Predictions with Machine Learning Hunt ...” Improving Earnings Prediction, 10/2019/12, https://zicklin.baruch.cuny.edu/wpcontent/uploads/sites/10/2019/12/Improving-Earnings-Predictions-with-MachineLearning-Hunt-Myers-Myers.pdf.